Datasets, that is, digital collections of related information, and data visualizations are really crucial ways for understanding the circulation of African American short stories and black literature in general. Calculating the number of unique stories and the many reprints allow us to survey how editors collectively shaped canonical histories. While conventional analog approaches such as standard bibliographies have and remain important, we can further advance African American literary studies by taking fuller advantage of digital tools and datasets, which greatly assist in exploring and quantifying multifaceted publishing histories.

This project developed from one central question: What short story writers did editors choose to republish most frequently? The search for answers involved me identifying 100 anthologies and tabulating information about the various collections. Anthologies often organize entries chronologically, but a dataset offers possibilities for arrangement by date, author birth year, author gender, entry title, and the original publication of a story. The dataset I created references short stories selected by more than 70 editors from 1925 to 2017.

The dataset identifies about 630 unique short stories by roughly 300 black writers. Those 630 unique stories have collectively been reprinted approximately 1,100 times. The way I managed and even became aware of all those numbers was through the creation of a series of datasets.

Over the course of the last 40 years in particular, anthology editors shaped the landscape of African American literature by repeatedly publishing approximately 30 stories by a core group of writers: Charles Chesnutt, Hurston, Richard Wright, Ralph Ellison, James Baldwin, Toni Cade Bambara, and Alice
Walker. Editors essentially established a “Big 7” of African American short fiction writers. The continual inclusion of these seven writers has solidified their presence in American and African American literature anthologies.

Drawing on the dataset I created of short story writers, I worked with UTA’s data librarian Peace Williamson to create a Tableau Public Visualization that highlights the significance of the Big 7, Langton Hughes, and other writers whose have at least four stories republished in anthologies. (Many short stories by Hughes appear in anthologies, but editors never come to a consensus about any specific story by him). The one-page visualization highlights the various contexts for the most frequently anthologized black writers and stories published in collections between 1925 and 2017. The visualization seeks to illuminate selected fields in the dataset thereby making information about 630 short stories and more than 1,100 reprints by 300 writers manageable.

African American publishing histories are multifaceted, and conventional bibliographies offer somewhat limited information, especially in comparison to many datasets. Thus, we should embrace the use of datasets in order to enhance our understanding of black cultural productions. In addition, data visualizations greatly assist us in comprehending the hundreds and sometimes thousands of bits of information that comprise a dataset.

I want to make it clear that my own developing projects are linked to other related projects. For one, my project echoes Maryemma Graham’s efforts to document African American novels with her Project on the History of Black Writing. Whereas she sought to document novels, I have focused on short stories. Furthermore, my project greatly benefited from my colleague Ken Roemer’s website Covers, Titles, and Tables: The Formation of American Literary Canons in Anthologies, which scanned the title pages of more than 100 literature collections. Like Graham and Roemer, I’m trying to grapple with large bodies of information concerning literary history. Datasets and data visualizations can facilitate what we can know and show.

And one final point, or a series of questions: What is the value of professional datasets on African American and American literature? That is, and more specifically, to what extent does a dataset uploaded to a public repository count toward tenure? I uploaded my “Black Short Story Dataset” to the Texas Data Repository. How far will that dataset count toward tenure? Would a search committee will that dataset as worth a full article, half an article, a quarter of article?

Raising these kinds of question about the worth of datasets will no doubt be essential moving forward. Only by placing more emphasis on datasets will assistant professors, for instance, be encouraged to devote quality time to constructing them.